

# Projection Back の性質に基づく分離行列の外挿を利用した音源分離\*

☆下川凜, 松本和樹, 吉野夏樹, 矢田部浩平 (農工大)

## 1 はじめに

優決定ブラインド音源分離 (優決定 BSS) 手法の一つである独立低ランク行列分析 (ILRMA) [1] は, 分離音のパワースペクトログラムを非負値行列因子分解 (NMF) に基づく低ランク行列でモデリングする. NMF の変数は通常, 乱数により初期化されるが [1], ILRMA の目的関数は非凸なため, 局所最適解に陥る場合がある. これを防ぐために, 一部の周波数帯域に ILRMA を適用し, その結果を用いて他の帯域を初期化する手法が提案されているが, その初期値の生成法にはランダム性が残っている [2]. 本稿では, 乱数を用いずに初期化した ILRMA で一部の帯域を分離したのちに, Projection Back (PB) の性質 [3] を利用して分離行列を外挿する. これにより, ランダム性のない ILRMA の初期化法を提案する.

## 2 ILRMA による優決定 BSS

優決定 BSS は, 時間周波数領域における  $M$  チャネルの観測信号  $\mathbf{x}_{t,f} \in \mathbb{C}^M$  をもとに, 各周波数  $f = 1, \dots, F$  ごとに分離行列  $\mathbf{W}_f \in \mathbb{C}^{N \times M}$  を推定することで,  $N$  ( $M \geq N$ ) 個の分離音を  $\mathbf{y}_{t,f} = \mathbf{W}_f \mathbf{x}_{t,f} \in \mathbb{C}^N$  の形で得る. ただし,  $t = 1, \dots, T$  は時間フレームのインデックスである. 優決定 BSS の代表的な手法である ILRMA は,  $n$  番目の分離音のパワースペクトログラム  $\mathbf{P}_n = (|y_{n,t,f}|^2)_{f=1,t=1}^{F,T} \in \mathbb{R}_+^{F \times T}$  を基底数  $K$  の基底行列  $\mathbf{T}_n \in \mathbb{R}_+^{F \times K}$  とアクティベーション行列  $\mathbf{V}_n \in \mathbb{R}_+^{K \times T}$  の積  $\mathbf{T}_n \mathbf{V}_n$  によりモデル化する. ただし, ILRMA の目的関数は非凸なため, 安定した分離のためには, NMF の変数  $\mathbf{T}_n, \mathbf{V}_n$  の適切な初期化法が求められる.

## 3 提案法

本稿では  $M = 2$  を想定し, PB の性質を利用して分離行列を外挿することで, 乱数を用いずに ILRMA を初期化する手法を提案する. 全体像を図-1 に示す.

### 3.1 一部の帯域に対する ILRMA ( $K = 1$ ) の適用

提案法では, 帯域を制限して分離をすることが安定性の向上に有効であり [2], また, パワースペクトログラムは狭い帯域ではランク 1 ( $K = 1$ ) の行列でモデル化しやすくなることに着目する. まずは観測信号のうち, 一

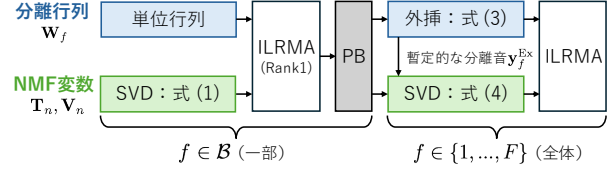


図-1 提案法の全体像.

部の帯域  $\mathcal{B} = \{f \mid f_{\min} \leq f \leq f_{\max}\}$  のみを, 特異値分解 (SVD) を用いて初期化した  $K = 1$  の ILRMA により分離する. 具体的には,  $\mathcal{B}$  内部の観測信号のパワースペクトログラム  $\mathbf{P}_n^{\mathcal{B}} = (|x_{n,t,f}|^2)_{t=1}^T)_{f \in \mathcal{B}} \in \mathbb{R}_+^{|\mathcal{B}| \times T}$  の SVD を  $\mathbf{P}_n^{\mathcal{B}} = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{Q}_n^T$  とし, その最大特異値に対応する成分を  $(\mathbf{U}_n)_{:,1} \in \mathbb{R}_+^{|\mathcal{B}| \times 1}$ ,  $(\mathbf{\Sigma}_n)_{1,1} \in \mathbb{R}_+$ ,  $(\mathbf{Q}_n)_{:,1} \in \mathbb{R}_+^{T \times 1}$  とする. ただし,  $(\cdot)_{:,k}$  は行列の  $k$  列目を取り出す操作を表す. これらを用いて, 観測信号の帯域  $\mathcal{B}$  のパワースペクトログラム  $\mathbf{P}_n^{\mathcal{B}}$  を最良ランク 1 近似する NMF の変数を

$$\begin{aligned} \mathbf{T}_n &= \max \left( \sqrt{(\mathbf{\Sigma}_n)_{1,1}} \cdot (\mathbf{U}_n)_{:,1}, \epsilon \right) \in \mathbb{R}_+^{|\mathcal{B}| \times 1} \\ \mathbf{V}_n &= \max \left( \sqrt{(\mathbf{\Sigma}_n)_{1,1}} \cdot (\mathbf{Q}_n)_{:,1}^T, \epsilon \right) \in \mathbb{R}_+^{1 \times T} \end{aligned} \quad (1)$$

で得る.  $\epsilon > 0$  は微小な定数である. また, 分離行列  $\mathbf{W}_f$  ( $f \in \mathcal{B}$ ) は単位行列で初期化する. これらの初期値を用いた ILRMA を, 観測信号のうち帯域  $\mathcal{B}$  を取り出したもの (i.e.,  $(\mathbf{x}_{t,f})_{t=1}^T)_{f \in \mathcal{B}}$ ) に対して適用し, 帯域  $\mathcal{B}$  の分離行列  $(\mathbf{W}_f)_{f \in \mathcal{B}}$  を最適化する.

### 3.2 全帯域に渡る ILRMA の初期値の構成

続いて, 3.1 節で得た帯域  $\mathcal{B}$  の分離行列をもとに, PB の性質を利用して全帯域の分離行列を外挿し, 暫定的な分離音を得る. この暫定的な分離音は観測信号と比較すると分離音に近い状態であると期待できる. したがって, そのパワースペクトログラムを近似する  $\mathbf{T}_n, \mathbf{V}_n$  を ILRMA の初期値として用いることで, 予め分離状態に近い初期値から最適化を開始でき, 安定した分離が可能であると考えられる.

具体的には, 帯域  $\mathcal{B}$  における分離行列  $\mathbf{W}_f$  にマイク 1 を基準とした PB が適用されているものと仮定すると, これらは振幅  $a_{i,f} \in \mathbb{R}_+$  と位相  $\phi_{i,f} \in \mathbb{R}$  (ただし,  $i = 1, 2$  かつ  $f \in \mathcal{B}$ ) を用いて

$$\mathbf{W}_f^{\text{PB}} = \frac{1}{D_f} \begin{bmatrix} a_{2,f} e^{j\phi_{2,f}} & -1 \\ -a_{1,f} e^{j\phi_{1,f}} & 1 \end{bmatrix} \quad (2)$$

\*Blind source separation using demixing matrix extrapolation based on the property of projection back. By Rin SHIMOKAWA, Kazuki MATSUMOTO, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

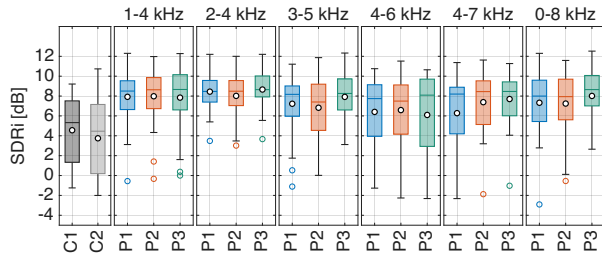


図-2 基底数 16 における帯域ごとの性能比較.

と書ける. ただし,  $D_f = -a_{1,f} e^{j\phi_{1,f}} + a_{2,f} e^{j\phi_{2,f}}$  である [3]. ここで, 式 (2) の右辺の行列のうち, 1 列目の 2 つの成分の位相成分を並べたベクトルを  $\mathbf{q}_i = (e^{j\phi_{i,f}})_{f \in \mathcal{B}} \in \mathbb{C}^{|\mathcal{B}|}$  とする. 提案法では,  $\mathbf{q}_i$  は複素正弦波を用いて近似できることに着目し, これを全帯域  $f = 1, \dots, F$  に渡って外挿する.

外挿の具体的な手順を述べる. まず,  $\mathbf{q}_i \in \mathbb{C}^{|\mathcal{B}|}$  に零ベクトルを連結したベクトル  $\mathbf{p}_i \in \mathbb{C}^F$  を作成する<sup>1</sup>. これをフーリエ変換することで, 最小二乗法に基づき主要な複素正弦波成分を取り出す. ここで,  $Q$  倍のオーバーサンプリングを伴う離散フーリエ変換  $\mathcal{F}_Q$  を, 位相の情報を持つベクトル  $\mathbf{p}_i$  に対して適用することで, ベクトル  $\mathbf{c}_i = \mathcal{F}_Q(\mathbf{p}_i) \in \mathbb{C}^{QF}$  を得る. その絶対値が最大となるインデックスを  $p_i^{\text{peak}} = \arg \max_{p \in \{1, \dots, QF\}} |(\mathbf{c}_i)_p|$  とする. 得られたピーク位置  $p_i^{\text{peak}}$  及び対応する位相成分  $\theta_i = \angle(\mathbf{c}_i)_{p_i^{\text{peak}}}$  から遅延量を求め, 外挿位相  $\hat{\phi}_i \in \mathbb{R}^F$  を  $\hat{\phi}_{i,f} = \theta_i + 2\pi p_i^{\text{peak}} f / QF$  ( $f = 1, \dots, F$ ) で求める. これを用いて, スルビームフォーマを

$$\mathbf{W}_f^{\text{Ex}} = \begin{bmatrix} 1 & -e^{-j\hat{\phi}_{2,f}} \\ -e^{j\hat{\phi}_{1,f}} & 1 \end{bmatrix} \quad (3)$$

で構成し, これを ILRMA における分離行列の初期値とし, 暫定的な分離音  $\mathbf{y}_{t,f}^{\text{Ex}} = \mathbf{W}_f^{\text{Ex}} \mathbf{x}_{t,f}$  を得る. そのパワースペクトログラム  $\mathbf{P}_n = (|y_{n,t,f}^{\text{Ex}}|^2)_{f=1,t=1}^{F,T}$  の SVD を  $\mathbf{P}_n = \mathbf{U}_n \mathbf{\Sigma}_n \mathbf{Q}_n^T$  とすると, 3.1 節と同様に,  $1, \dots, K$  までの特異値に対応する成分を用い, NMF の変数の初期値を以下の式で得る.

$$\begin{aligned} \mathbf{T}_n &= \left( \max \left( \sqrt{(\mathbf{\Sigma}_n)_{k,k}} \cdot (\mathbf{U}_n)_{:,k}, \epsilon \right) \right)_{k=1}^K \\ \mathbf{V}_n &= \left( \max \left( \sqrt{(\mathbf{\Sigma}_n)_{k,k}} \cdot (\mathbf{Q}_n)_{:,k}^T, \epsilon \right) \right)_{k=1}^K \end{aligned} \quad (4)$$

## 4 実験

SiSEC2011 dev1 データセットより作成した 2 音源 2 チャンルの混合音 24 組を用いて分離実験を行った. 比較手法一覧を表-1 に示す. 従来手法として, 初期化に単位行列と乱数を用いる通常の ILRMA を Conv1, 式 (4) を利用して観測信号から NMF の変数

<sup>1</sup>この時,  $(\mathbf{p}_i)_f = q_{i,f}$  ( $f \in \mathcal{B}$ ) かつ  $(\mathbf{p}_i)_f = 0$  ( $f \notin \mathcal{B}$ ).

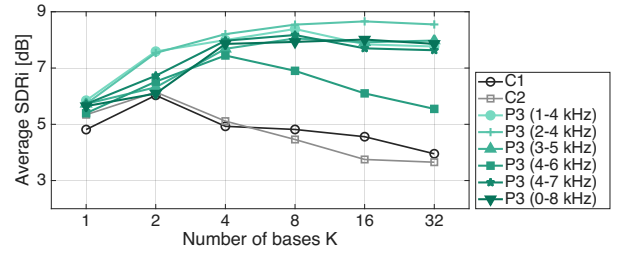


図-3 基底数ごとの性能の変化.

表-1 初期化手法の組み合わせ

| NMF の変数      | 乱数         | 観測信号       | 暫定分離音の SVD 式 (4) |
|--------------|------------|------------|------------------|
| 分離行列         |            |            |                  |
| 単位行列         | Conv1 (C1) | Conv2 (C2) | Prop1 (P1)       |
| 外挿分離行列 式 (3) | Prop2 (P2) | —          | Prop3 (P3)       |

を初期化する手法を Conv2 とした. 提案法は式 (3), (4) の各初期化の有無により Prop1-3 とした. なお, NMF の変数を乱数で初期化した Conv1, Prop2 については, 乱数を用いるため, 5 回の平均性能とした. 前処理の白色化は省略し, 後処理として PB を用いた. 反復回数は従来手法を 120 回, 提案法では一部の帯域を 20 回, その後の全帯域を 100 回とした. STFT の窓長は 2048, シフト量は 1024 とし, オーバーサンプリング  $Q$  は 64 倍とした. 評価には SDRi を用いた. 帯域  $\mathcal{B}$  (1-4, 2-4, 3-5, 4-6, 4-7, 0-8 kHz) および基底数 (1, 2, 4, 8, 16, 32) を変化させて評価した.

基底数が 16 のときの実験結果を図-2 に示す. 提案法は, 2-4 kHz, 4-7 kHz など適切な帯域を選ぶことで分離が安定した. また, 提案する分離行列, 及び NMF の変数の初期化法は, とともに分離性能の向上に寄与した. 基底数ごとの性能を図-3 に示す. 通常の ILRMA は,  $K = 2$  のとき, もっとも高い性能を達成した一方, 提案法は多くの場合, 基底数を増やすことで性能が向上する傾向が確認できた.

## 5 むすび

本稿では, 一部帯域の分離結果を用いたランダム性のない ILRMA の初期化手法を提案し, その有効性を確認した. 今後は, 初期化に用いる帯域の検討や, 十分な分離性能が得られなかった音源などに注目し, 初期化手法の改善を図る.

## 参考文献

- [1] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **24**(9), 1626–1641 (2016).
- [2] K. Matsumoto and K. Yatabe, “Subband splitting: Simple, efficient and effective technique for solving block permutation problem in determined blind source separation,” *Acoust. Sci. & Tech.*, **46**(4), 326–337 (2025).
- [3] 下川凜, 松本和樹, 矢田部浩平, “ブラインド音源分離における  $2 \times 2$  分離行列の自由度について,” 日本音響学会講演論文集, pp. 261–262 (2026.3).